DOCUMENT RESUME

ED 431 950 CE 078 982

TITLE	Evaluation in Human Resource Development.
PUB DATE	1999-03-00
NOTE	29p.; In: Academy of Human Resource Development (AHRD) Conference Proceedings; see CE 078 963.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Accountability; Adult Education; Comparative Analysis; Computer Assisted Testing; Distance Education; Education Work Relationship; Educational Research; Employee Attitudes; Evaluation Methods; Foreign Countries; Higher Education; Interactive Video; Job Performance; Job Satisfaction; Job Training; *Labor Force Development; Online Systems; *Organizational Development; Outcomes of Education; Performance; Program Effectiveness; *Program Evaluation; Test Construction; Transfer of Training
IDENTIFIERS	Finland

ABSTRACT

These four papers are from a symposium on evaluation in human resource development (HRD). "Assessing Organizational Readiness for Learning through Evaluative Inquiry" (Hallie Preskill, Rosalie T. Torres) reviews how evaluative inquiry can facilitate organizational learning; argues HRD evaluation should be reconceptualized as a process for assisting learning; and describes efforts to develop an instrument HRD professionals can use to assess an organization's readiness to engage in learning through evaluative inquiry. "The Design and Development of an Evaluation System for Online Instruction" (Tim L. Wentling, Scott D. Johnson) describes development of a system using computer technology to provide partial automation. "The Mixed-Method Evaluation in HRD--A Description of Methods Applied in the Industrial Development Program" (Ville Nurmi) describes a study in the Finnish paper industry that concluded that mixed-methods is a purposeful design when the aim is to assess a large program from multiple viewpoints. "The Effect of Two Training Systems on Trainee Satisfaction, Learning, Job Behavior, Retention and Organizational Performance" (Maria Hruby Moore) compares a manager-led, self-paced videotape delivery system and an instructor-led group method over two-way interactive videoconferencing and finds no significant difference between trainee satisfaction, learning, and individual job performance outcomes but significant differences between learning ratings, retention rate, and organizational performance measures. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 431 950

18 **Evaluation in Human Resource Development**

Chair: Darlene Russ-Eft, AchieveGlobal, Inc.

18 - 1 Assessing Organizational Readiness for Learning through Evaluative Inquiry

Hallie Preskill, University of New Mexico

Rosalie T. Torres, Developmental Studies Center

18 - 2 Design and Development of an Evaluation System for Online Instruction

Tim L. Wentling, University of Illinois at Urbana/Champaign

Scott D. Johnson, University of Illinois at Urbana/Champaign

18 - 3 The Mixed-Method Evaluation in HRD--A Description of Methods Applied in the Industrial Development Program

Ville Nurmi, Finnish Institute of Occupational Health

18 - 4 The Effects of Two Training Systems on Trainee Satisfaction, Learning, Job Behavior, Retention and Organizational Performance

Maria Hruby-Moore, Ohio State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

H Preskill V Nurmi
T Wentling M T Moore

2A

CE 078582

BEST COPY AVAILABLE

Assessing Organizational Readiness for Learning Through Evaluative Inquiry

Hallie Preskill
University of New Mexico
Albuquerque, New Mexico

Rosalie T. Torres
Developmental Studies Center
Oakland, California

In this paper we first review the concept of organizational learning and how it can be facilitated by evaluative inquiry. We then describe the status of HRD evaluation and argue that HRD evaluation should be reconceptualized as a process for facilitating individual, team and organizational learning. Finally, we describe our efforts to develop an instrument that HRD professionals can use for assessing an organization's readiness to engage in learning through evaluative inquiry.

Keywords: Evaluation, Organizational Learning, Assessment

To say that times are changing seems trite today and terribly obvious. In just a few years we have seen major shifts in how organizations are meeting the challenges of a changing global economy, emerging technologies, and an evolving, ever-expanding, multi-cultural workforce. While writers on organizational change and management theory offer a plethora of recipes for achieving success in this dynamic environment, there is growing consensus that the organizations which survive will be flexible, adaptive, and quick to respond to changing conditions, and will use information to make decisions. These organizations understand the value of human capital and will develop systems and processes that maximize all employees' ability to contribute to the organization's success. Intellectual capital -- employees' knowledge, information, intellectual property, and experience will determine the organization's future (Stewart, 1997). As Davis (1996) explains, organizations that make the transition to the knowledge era will learn how to generate, use, and sell information, regardless of the business they are in.

Organizational Learning

Even before management experts started talking about intellectual capital and the coming knowledge era, Argyris and Schon (1978, 1996) were exploring the concept of organizational learning and its role in facilitating organizational change. Since then, many other researchers have contributed to our understanding of organizational learning principles and the characteristics of learning organizations (for example, see DiBella & Nevis, 1997; Dixon, 1994; Marquardt, 1996; Senge, 1990; Watkins & Marsick, 1993). While some researchers and theorists have focused on how organizations as whole systems adapt or change, others have focused on how individuals learn in organizations; that is, how they develop, adapt and update their cognitive models (Edmondson & Moingeon, 1998). However, based on their review of 22 case examples of organizational learning, Watkins and Marsick (1996, p. 10) conclude that most models share the following attributes:

- Learning organizations focus on organizational learning and transformation; it is not enough for individuals to learn.
- Structures and systems are created to ensure that knowledge is captured and shared for use in the organization's memory.
- Leaders and employees at all levels think systematically about the impact of their decisions and work within the total system.
- Learning is built into work structures, politics, and practices.
- Learning is transformative in some way, although it is likely that some new learning will also be adaptive.
- Learning has a greater impact when it involves a greater percentage of the employee population.
- Organizational systems and politics are structured to support, facilitate, and reward learning for individuals, teams, and the organization.

Copyright © 1999 Hallie Preskill & Rosalie T. Torres

- Measurement systems benchmark current knowledge and culture and monitor progress toward becoming a learning organization.

In sum, organizational learning can be viewed as, "a continuous process of organizational growth and improvement that: (a) is integrated with work activities, (b) invokes the alignment of values, attitudes, and perceptions among organizational members, and (c) uses information or feedback about both processes and outcomes to make changes" (Torres, Preskill & Piontek, 1996, p. 2). In this sense, organizational learning represents the organization's commitment to using all of its members' capabilities. It is different from individual and team learning in that organizational learning is dependent on individuals and teams sharing their learning in an ongoing, systemic way. Ultimately, organizational learning is about creating continuous processes and mechanisms for learning about how to do things better. In the next section, we discuss the relationship among traditional evaluation practice in HRD, evaluative inquiry, and organizational learning.

Linking Organizational Learning and Evaluative Inquiry

The term evaluation has traditionally been defined as a process of gathering information to make determinations about a program's value, merit or worth (Scriven, 1991 Patton, 1997). In the evaluation field there has been a recent shift from viewing evaluation as a summative activity conducted by external evaluators, to one that is more participatory, internal, and improvement-focused (Preskill & Caracelli, 1997). In the HRD field, evaluation has often been considered part of a program's design process (Andrews & Goodson, 1980; Banathy, 1968; Dick & Carey, 1985; Rothwell & Kazanas, 1998). In linear instructional design (ISD) models, evaluation is typically placed at the end-point of the model. For this reason, and for others, evaluation has been construed as a costly and time-consuming add-on activity that occurs more as an event than as a process (Hilbert, Preskill & Russ-Eft, 1997). As a result, the benefits of evaluation are often overlooked or ignored as a means for providing relevant and useful information for decision-making.

Given the changing nature of organizations and societies, we have come to believe that evaluation must also be reconceptualized to focus on how individuals, team, and organizations can learning from evaluative inquiry (Preskill & Torres, 1999). As a result of our research over the last five years, we have developed a framework for helping organizations engage in evaluative inquiry that results in individual, team and organizational learning. Our framework is based on the following assertions:

- Dialogue, reflection, asking questions and identifying and clarifying values, beliefs, assumptions and knowledge are critical to learning from evaluative inquiry.
- Learning occurs through the social construction of knowledge and can be transformative when stakeholders are able to alter their perceptions and understandings of that which is being evaluated.
- Learning from evaluation occurs within the context of the organization and is therefore mediated by the organization's internal systems and structures.
- Evaluation must be increasingly responsive to the evolving information and decision- making needs of organizations.
- Evaluative inquiry should be ongoing and integrated into all work practices.

Evaluative inquiry for organizational learning and change encompasses the following:

- A focus on program and organizational processes as well as outcomes;
- Shared individual, team and organizational learning;
- Education and training of organizational practitioners in inquiry skills;
- Modeling the behaviors of collaboration, cooperation, and participation;
- Establishing linkages between learning and performance;
- Searching for ways to create greater understanding of the variables that affect organizational success and failure; and,
- Using a diversity of perspectives to develop understanding about organizational issues.

While evaluation's primary focus has been on the "program" as the unit of analysis, evaluative inquiry addresses issues and concerns individuals may have about various processes and systems at the departmental or

organizational level as well. Thus, the unit of analysis is wherever the problem or concern is identified (Preskill & Torres, in press).

Phases of Evaluative Inquiry for Learning in Organizations

Evaluative inquiry for learning consists of three specific phases (see Figure 1). In the *Focusing the Inquiry* phase, evaluation team members determine what issues and concerns the evaluative effort will address, who the stakeholders are, and what questions will guide the evaluative inquiry. In the next phase, *Carrying Out the Inquiry*, organization members determine the most appropriate inquiry design, methods of data collection, analysis and interpretation, and communicating and reporting strategies. The third phase, *Applying Learning* , asks organization members to develop strategies that address the evaluative inquiry's outcomes, design and implement action plans based on these strategies, and monitor the progress of actions taken.

Evaluative Inquiry Learning Processes

As each of the inquiry phases (Focusing the Inquiry, Carrying Out the Inquiry, and Applying Learning) are implemented, organization members come together to engage in the learning processes of: (1) Dialogue, (2) Reflection, (3) Asking Questions, and (4) Identifying and Clarifying Values, Beliefs, Assumptions, and Knowledge (see Figure 1). By participating in these processes throughout evaluative inquiry efforts, greater insights and understandings about organizational issues are developed. Ultimately, these insights and understandings lead to informed decisions for organizational change. It is important to note that we do not view these processes as linear. For example, individuals do not necessarily first reflect, then ask questions and then identify their own values and beliefs. Rather, these learning processes occur through a dynamic, fluid, social interaction among organizational members.

Elements of Organizational Infrastructure

In large part, the success of evaluative inquiry is based on an organization's infrastructure; that is, the strength of the underlying foundation or framework for supporting learning within the organization. An organization's infrastructure can strongly influence the extent to which organization members learn from evaluative inquiry and use their learning to support personal and organizational goals. The elements of the organization's infrastructure include: culture, leadership, communication, systems and structures. The nature of these components provides the footing on which evaluative inquiry efforts can be undertaken and sustained. That is, these components will facilitate or inhibit organizational learning from evaluative inquiry to varying degrees, depending on how they operate within the organization.

The Need for a Readiness Instrument to Assess Learning Through Evaluative Inquiry

As we've worked with organizations interested in applying our framework, we've learned that organizations are often uncertain of where to start. They are unclear as to where their organization currently stands with regard to their readiness not only for organizational learning, but also for learning from evaluation, in particular. Most organizations have little experience with evaluation and even less with linking evaluation to organizational learning concepts and practices.

Assessing Organizational Learning

While the last twenty years have been devoted to developing theories about organizational learning, several researchers and consultants have recently begun developing related assessment instruments. The majority of these instruments are designed as diagnostic tools to determine the extent to which an organization possesses the characteristics of a learning organization (Redding, 1997). The questionnaires tend to ask individuals to think about their department/unit or organization as a whole, and to respond to items regarding the extent to which learning practices are supported at the individual, team and/or organizational levels. Specific dimensions, domains, subsystems, or practices that the authors believe are important to organizational learning generally form the item categories. These categories include vision, leadership/executive practices, culture/climate, organizational structure,

human resource systems, individual and team learning/practices, and work processes. The results of these assessments are used to help an organization improve its organizational learning capabilities. Though Yang, Watkins and Marsick (1998) suggest that "constructing a valid instrument is an on-going process," their preliminary results are showing an "underlying structure that represents patterns of learning activities in the organization"(p.3-3).

Assessing an Organization's Readiness for Evaluation

As organizations increasingly recognize the need to better understand how they can improve their systems and processes to compete in the new techno-global environment, they are expressing an interest in evaluating the effectiveness and impact of human resource development (HRD) efforts (Bassi, Benson & Cheney, 1996; Industry Report, 1996). However, most organizations are uncertain about how to conduct rigorous, useful evaluations or how to find the resources and time to do them well (ASTD, 1997). Few trainers have any academic background in evaluation, or much practical experience conducting systematic evaluations. In addition, the models, frameworks, and taxonomies for conducting evaluation that are prevalent in the HRD scholarly and practitioner literature provide little information about the context, procedures, and issues inherent in conducting program evaluations (Hilbert, et al., 1997). Practitioners who understand the value of evaluative inquiry for addressing programmatic and organizational issues worry that their organizations are not "ready" to evaluate; that the organization's culture doesn't appear to support evaluation, or that there are few mechanisms to support evaluation activities. Yet, there remains a growing need and interest in how to effectively evaluate how well and in what ways, HRD programs are contributing to the organization's goals and mission.

In spite of evaluation being a topic in the HRD literature for the last 40 years, as well as 30 years of research on evaluation in the evaluation field, we were only able to locate one instrument on assessing an organization's readiness for evaluation (Mayer, 1976). In some ways, this finding is not terribly surprising. In the HRD field, evaluation has rarely been viewed as a strategic organizational activity. It has traditionally been treated as an after-thought or as a post-course reaction survey tacked on to the end of a training session. There is scant literature on the role evaluation can play in helping organizations achieve their goals, or how HRD evaluation can facilitate the organization's growth and development (Hilbert, et al.).

In terms of the evaluation field, where one might expect to see more interest in developing such diagnostic instruments, there are other possible explanations for its neglect. The field itself, and much of what we know about evaluation today, grew out of the social program legislation in the 1960's and has a history of focusing on large-scale social and educational programs (see House, 1993; Shadish, Cook & Leviton, 1995). Using an assessment instrument to measure an organization's readiness for evaluation would not have been seen as applicable or appropriate in many situations. However, given that organizations in every sector are recognizing the need for evaluating programs at an organizational level, it is perhaps timely to develop such measures. The need for a readiness for evaluation instrument is based on the idea that it would be a waste of resources to engage in evaluative inquiry if the organization is not ready, interested, or willing to use the results of an evaluation. Therefore, such an instrument would yield information that would help prepare and guide the organization in their efforts to evaluate various programs, processes, products, and services.

Developing an Instrument to Assess Learning from Evaluative Inquiry

In response to these needs, we are developing an instrument to assess an organization's readiness to learn from evaluative inquiry. While organizational learning assessment instruments do exist, and are being validated, as indicated earlier, none make any kind of link between evaluation and organizational learning. Thus, we believe there is a need for another instrument that specifically measures an organization's readiness to learn from evaluative inquiry.

By developing and using this instrument within organizations, organizations will be better prepared to use evaluative inquiry practices to address issues that challenge their success. As a result, organizations may benefit in a number of ways. Individuals and team members may: (a) better understand how their actions affect other areas of the organization, (b) ask more questions than give solutions/answers, (c) develop a greater sense of personal accountability and responsibility for the organization's outcomes, (d) act more consultative, (e) be more likely to ask for help, (f) use information to act, and (g) be more willing to share the work that needs to be done. These outcomes may result in an organization's enhanced ability to: (a) develop new products and services, (b) increase productivity, (c) improve morale, (d) experience less turnover, (e) experience less waste/sabotage/error, (f) experience improved

financial performance, (g) experience increased efficiency and less redundancy, (h) provide more effective service to clients/customers, and (i) change more quickly. (Preskill & Torres, 1999)

Research Questions

The following questions are guiding the development and validation of the readiness for learning through evaluative inquiry instrument:

1. What are the questions/items that effectively measure readiness for evaluative inquiry and organizational learning?
2. What is the content and construct validity of the assessment instrument?
3. What is the reliability of the assessment instrument?
4. How can the instrument be used within organizations?

Methodology

The instrument's development began by generating items from three sources: (a) a review of the organizational learning and program evaluation literatures, (b) a review of existing learning organization/organizational learning and readiness for evaluation instruments, and (c) interviews conducted with four organizations that are attempting to integrate organizational learning practices and systems within their organizations. This process first yielded 164 items. The items were then classified into twelve categories, or dimensions based on the *Evaluative Inquiry for Learning in Organizations* framework (Preskill & Torres, 1999).

Nine HRD graduate students were invited to participate in a modified Q-sort process. Each of the 164 items and dimension titles were printed onto cardstock and randomly placed in a stack. Participants were asked to read each potential item and to place it under one of the twelve dimensions. After this process was completed, each of the participant's classifications was matched to the authors' initial classification system. As a result, items that were sorted into different categories were reviewed and resorted by the authors. Based on the results of the modified Q-sort, the number of dimensions was reduced to six, and the items to 82.

The next step is to again have a group of graduate students classify the 82 items into the six dimensions. Based on their sorts, additional revisions will be made. In is anticipated, however, that the final version of the instrument will contain between 70-85 items in addition to some demographic items. The purpose of this activity will be to collect evidence of the items' content validity.

Following the second Q-sort, the next set of tasks will involve establishing the instrument's internal reliability and construct validity. Our intent is to pilot the instrument with approximately 200-300 individuals from various organizations. Item analyses will be conducted on the pilot data to determine the internal consistency of the dimensions established on the basis of the Q-sort data. Items with low discriminatory power will be reviewed for revision or deletion. To determine construct validity, we will examine the results of confirmatory factor analysis (CFA). This technique will be used to verify the adequacy of each item-to-factor association and the number of dimensions underlying the construct.

Results and Conclusions

Given the early phase of this research, we are unable at this time to report on the results of the pilot test (research questions 2 and 3). However, we believe that the instrument will provide HRD researchers, consultants, and practitioners with valuable information as they strive to integrate evaluative inquiry into an organization's learning practices. We see evaluative inquiry as a useful approach to helping organization members answer pressing questions that challenge their success and future. In so doing, it can provide critical information for making decisions that result in increased individual, team and organizational learning.

By developing and validating this readiness instrument, HRD professionals will be in a better position to help organizations know where their learning and evaluation strengths lie, and provide them with ideas and strategies about how to leverage these strengths in ways that maximize learning from evaluative inquiry activities. The readiness instrument will be a tool to begin the process of building evaluative inquiry into the learning systems and practices of organizations.

References

Andrews, D. & Goodson, L. "A comparative analysis of models of instructional design. Journal of Development, 3 (4), 2-16.

ASTD National HRD Executive Survey (1997). Measurement and evaluation, 1997 Fourth Quarter Survey Report. Alexandria, VA: American Society for Training and Development.

Banathy, B. H. (1968). Instructional systems. Palo Alto, CA: Fearon Publishers.

Bassi, L. J., Benson, G. & Cheney, S. (1996). The top ten trends. Training & Development, 50 (11), 28-42.

Davis, S. (1996). Rumble, rumble. Training & Development, 50 (11), 44-45.

DiBella, A. J. & Nevis, E. C. (1997). How organizations learn. San Francisco, CA: Jossey-Bass, Inc.

Dick, W. & Carey, L. (1985). The systematic design of instruction. Second edition. Glenview, IL: Scott, Foresman and Company.

Dixon, N. (1994). The organizational learning cycle. London: McGraw-Hill.

Edmondson, A. & Moingeon, B. (1998). From organizational learning to the learning organization. Management Learning, 29 (1), 5-20.

Hilbert, J., Preskill, H. & Russ-Eft, D. (1997). Evaluating training's effectiveness. In L. Bassi & D. Russ-Eft (Eds.), What works: Assessment, development and measurement. Alexandria, VA: ASTD.

House, E. R. (1993). Professional evaluation: Social impact and political consequences. Thousand Oaks, CA: Sage.

Industry Report . (October, 1996) Training Magazine, 33 (10), 37-79.

Mayer, S. E. (1996). Organizational readiness to accept program evaluation questionnaire. Program Evaluation Resource Center, Minneapolis, MN.

Patton, M. Q. (1997). Utilization-focused evaluation: New century text. 3rd Edition. Thousand Oaks, CA: Sage.

Preskill, H. & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use TIG survey results. Evaluation Practice, 18 (3), 209-225.

Preskill, H. & Torres, R. T. (1999). Evaluative inquiry for learning in organizations. Thousand Oaks, CA: Sage.

Preskill, H. & Torres, R. T. (in press). Building capacity for organizational learning through evaluative inquiry. Evaluation.

Redding, J. (1997). Hardwiring the learning organization. Training and Development, 51 (8),

Rothwell, W. J. & Kazanas, H. C. (1998). Mastering the instructional design process. 2nd Edition. San Francisco, CA: Jossey-Bass.

Scriven, M. (1991) Evaluation thesaurus. Fourth edition. Thousand Oaks, CA: Sage.

Senge, P. M. (1990). The fifth discipline. New York: Doubleday.

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). Foundations of program evaluation: Theories of practice. Thousand Oaks, CA: Sage.

Stewart, T. A. (1997). Intellectual capital. New York: Doubleday/Currency.

Torres, R. T., Preskill, H. S. & Piontek, M. (1996). Evaluation strategies for communicating and reporting: Enhancing learning in organizations. Thousand Oaks, CA; Sage.

Watkins, K. E. & Marsick, V. J. (1993). Sculpting the learning organization. San Francisco: Jossey-Bass Publishers.

Watkins, K. E. & Marsick, V. J. (Eds.) (1996). Creating the learning organization. Volume 1. Alexandria, VA: ASTD

Yang, B., Watkins, K. E., & Marsick, V. J. (1998). Examining construct validity of dimensions of the Learning Organization Questionnaire. Academy of Human Resource Development (AHRD) Conference Proceedings. Paper presented at the AHRD annual conference, March, 1998.

The Design and Development of an Evaluation System for Online Instruction

Tim L. Wentling
Scott D. Johnson
University of Illinois at Urbana-Champaign

With the growth in popularity of online instruction has also come the concern for accountability and the need to make adjustments and improvements in online programs. This paper describes the conceptualization and development of an evaluation system that can be used to judge online instructional efforts. The evaluation system addresses concerns of both program administrators and course instructors. Computer technology is used to provide partial automation to reduce respondent burden and to efficiently use evaluation resources.

Keywords: Program Evaluation, Evaluation System, Online Instruction

Online or web-based learning is an education and training phenomena that is growing rapidly around the globe. As with any innovation, there are questions about effectiveness, efficiency, and utility (McCollum, 1998). With online instruction, the issue of evaluation and accreditation is especially important due to various threats commonly perceived by educators and trainers. These threats include a change in student campus life and the roles of professors (McCollum, 1998), the fear that the push for distance learning comes from entrepreneurs and university "bean counters" rather than educators (Mendels, 1999), and whether virtual students feel more isolated than their traditional counterparts (Arenson, 1998). These concerns can only be addressed through systematic research and evaluation efforts.

Human resource development professionals are blessed with a rich history of evaluation thought, development, and research, primarily derived from the fields of education (Alkin, 1969; Patton, 1997; Scriven, 1967; Stake, 1967, 1978, 1981; Stufflebeam, 1967, 1985; Stufflebeam, Foley, Gaphart, Guba, Hammond, Merriman, & Provus, 1971; 1997) and human resource development (Holton, 1995; Kirkpatrick, 1976; 1996; 1998; Kaufman & Keller, 1994). From these foundations, specialized evaluation processes and systems have been designed and implemented for HRD interventions, including training programs (Raab, Swanson, Wentling, & Clark, 1992, Wentling, 1980, 1992, 1993).

The design of evaluation systems for online instruction has been attempted by a number of professionals, including instructional designers of CBT systems (Clark, 1994; Draper, 1996; Simonson, 1997) technology experts (Jackson, 1990; Kimball, 1998; Middleton, 1997), and HRD practitioners (Magalhaes & Schiel, Pisik, 1997). The work of these individuals demonstrates progress in applying intuitive principles and practices of evaluation to online environments. Other authors have provided practical applications of evaluation theory by offering suggestions and guidelines for the evaluation of online instruction (Khan, 1997; Nichols, 1997; Oakes, 1997; Ravitz, 1997; Thorpe, 1993).

Problem Statement

The evaluation of online instruction is an important part of the design and implementation process. Even though practitioners and academics currently evaluate online instruction, their attempts have been limited to the use of traditional research methods and intuitive approaches to evaluation. There appears to be a lack of systematic evaluation of online programs that is built on evaluation theory and practice. The activity described in this paper is an attempt to conceptualize and develop a transferable, adaptive evaluation system for online instruction.

Conceptualization and Development of the Evaluation System

The purpose of the project was to design, develop, and implement an evaluation system that would meet the needs of the developers and sponsors of online instructional programs. The activities of the project reflect major efforts to obtain information, design components of the evaluation system, and test and revise the evaluation system. The specific activities of the project were to (1) develop a conceptual model for online evaluation, (2) identify specific

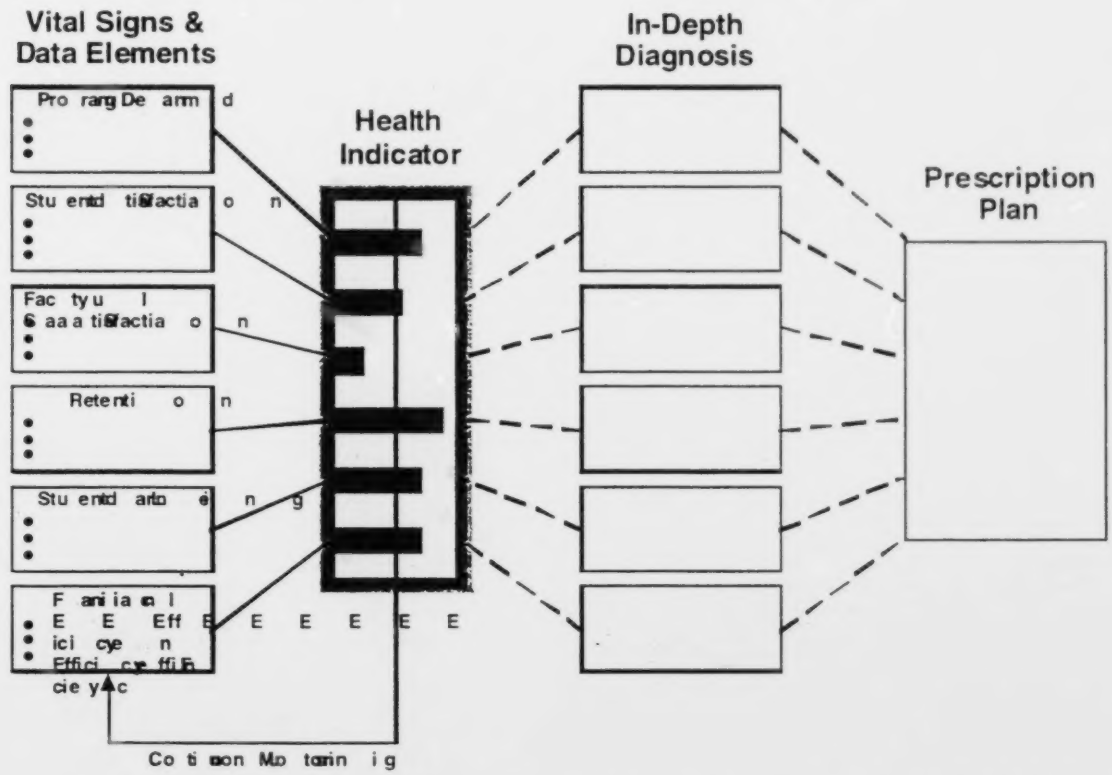
Copyright © 1999 Wentling & Johnson

vital signs of an online program and determine appropriate measures, (3) automate the collection and analysis of evaluation data through knowledge engineering approaches, (4) create an electronic performance support system to assist in program evaluation, and (5) conduct a field test of the evaluation system. The following sections describe each of the five major activities that are being accomplished in the formulation of the Illinois Online Evaluation System. At the time of this writing the evaluation system is evolving. Many of the tasks are complete, while some are underway.

Activity 1: Development of a Conceptual Model for Online Evaluation

This activity involved the conceptualization and design of the overall evaluation system, with the identification of major components and their functional relationships. The evaluation system, as currently designed, occurs in three stages: (1) vital sign assessment, (2) in-depth analysis, and (3) program improvement planning (see Figure 1). The evaluation system can best be explained using a medical analogy in which a physician examines a patient's vital signs to determine the patient's current state of health. Vital signs that are below acceptable standards are examined in more detail, utilizing more precise information and investigative techniques. As a result of the analysis, a plan is developed to facilitate improvement.

Figure 1. The Online Evaluation System.



Vital sign assessment (Stage 1) diagnoses the general health of the online program and individual online courses using data collected through routine activities. Example data include the number of inquiries about the program, the number of applications received, and the performance of students in the individual courses. The data are used to calculate a program's "health" rating in six areas: (1) student demand, (2) student retention, (3) student satisfaction, (4) faculty satisfaction, (5) student achievement, and (6) financial efficiency. The ratings help program personnel and sponsors monitor the overall performance of the program and individual courses to identify areas that may be problematic.

During the *in-depth analysis* (Stage 2), any vital signs that are identified as substandard are analyzed in detail. A thorough investigation is conducted to determine underlying problems or causes. Often this investigation

requires the collection of additional data, or the use of data from related vital signs. For example, a low rating in student satisfaction might lead to the further analysis of the individual items that comprise the student rating indicator. Additionally, a review of data related to student learning and faculty satisfaction might be done.

Program improvement planning (Stage 3) provides solutions to the problems that are investigated during the in-depth analysis. Alternative strategies for solving the problems are identified, along with resource requirements for implementing the solutions. Program personnel are able to select various courses of action and develop plans for addressing the identified problems.

Activity 2: Vital Sign Identification and Development of Measurement Instruments and Procedures

The major activity in the development of the evaluation system involved several inquiry endeavors to identify "quality indicators" for use as assessment criteria. Lists of quality indicators drawn from the literature on computer mediated education, outcomes of education, and evaluation were identified and prioritized. An initial list of 18 "vital signs," developed from the results of the literature was reviewed by a group consisting of HRD and evaluation experts. The list of vital signs was reduced to six by determining the relative importance of each to the stakeholders of the evaluation.

Instruments and procedures for gathering, summarizing, and analyzing vital sign data are being developed by project staff. Data used to calculate vital sign ratings are obtained through analysis of electronic archives, document review, surveys, student testing, and expert review. Sample data elements used to calculate each vital sign rating are described in Table 1. The actual vital sign ratings are calculated from student test scores, mean scores on survey instruments (i.e., student satisfaction and faculty satisfaction), and data comparisons (e.g., enrollment, retention, financial data). Standard transformations of the data are performed to provide a common scale for each vital sign to facilitate portrayal and comparison across the vital signs.

Table 1
Data Elements for Vital Sign Rating Calculation.

Vital Signs	Data Elements
Program Demand	8. Number of applications requested per semester
	9. Number of applications received per semester
	10. Number of telephone contacts per semester
Student Satisfaction	11. End of course student ratings of program content, quality of instruction, instructional resources, technology used, amount of interaction, instructional methods
	12. Mid-semester student ratings
	13. CISS data regarding perceptions of course interaction, course structure, and technical support
Faculty Satisfaction	14. Faculty ratings of technology, technical support, interaction with students, quality of student work
Student Retention	15. Percentage of dropouts from beginning to end
Student Learning	16. Self-assessment of learning
	17. Course project scores
	18. Quiz and Test scores
	19. Course grades
Financial Efficiency	20. Course design cost
	21. Course delivery cost
	22. Publicity cost
	23. Total unit cost, direct cost, overhead cost,
	24. Tuition revenue

Activity 3: Knowledge Engineering

As work progressed in conceptualizing the evaluation system and identifying vital signs, alternatives for automating the system were explored. Literature on artificial intelligence, decision support software, electronic performance support systems (EPSS), and expert systems was reviewed by project staff. Due to the technology potential for the evaluation model, its focus on internal rather than external evaluation, and its purpose as a self-evaluation tool to be used by a variety of online personnel, including program administrators, instructors, and instructional designers, it was ultimately decided to utilize performance improvement technologies to automate much of the data collection and analysis functions. Performance improvement technologies make dynamic use of technology to facilitate data collection and analysis tasks and to improve individual performance by providing timely information, advice, coaching, and training. An electronic performance support system is an example of a performance improvement technology that "captures, stores, and distributes individual and corporate knowledge assets throughout an organization to enable an individual to achieve a required level of performance in the fastest possible time and with the minimum of support from other people" (Raybould, (1995), p. 11). EPSS technologies often include an information database, an expert advisor, customized tools and templates, and the potential to run simulations. Beyond a few efforts in corporate settings, this technology has not been widely applied in education and training for evaluation purposes.

Available literature on performance improvement technologies was reviewed. A key component to any expert system is the knowledge base of a recognized expert in the problem area (Hayes-Roth, et al., 1983). Focus groups of recognized online education experts from the University of Illinois were held to ascertain their perceptions of what constitutes effective online education and evaluation. These experts worked through simulated evaluation problems to test their ideas and were asked to develop specific guidelines and suggestions for improving the vital sign assessment and in-depth procedures. The framework for the expert system was based on the outcomes of this activity.

Activity 4: EPSS Development

The investigative framework gleaned from the knowledge engineering activity was developed into a series of If-Then rule process charts (one for each vital sign) by the developers. One partial example of a vital sign rule process chart is shown in Table 2 where a variety of "then" statements are provided in response to the low indication of student satisfaction.

Table 1
Sample EPSS Rule Process Chart for One Vital Sign

Vital Signs	Data Elements
IF: Student Satisfaction as indicated by Student Rating Form is Low,	THEN: 4. review subscores for student ratings results. 5. identify the subscores that reveal the area of problem. 6. review results to individual items in the problem subscore(s).
IF: If more information is needed to determine the problem,	THEN: 7. develop an interview questionnaire and contact a sample of students by phone. 8. develop an e-mail questionnaire with relevant questions and send to a sample of students. 9. summarize new student data. 10. present the summary of all results to the instructor and discuss potential causes of the problem.

The probable causes identified by the online instruction stakeholders were incorporated into the rule process charts to complete the exploration diagrams. Sample computer display screens were also developed that simulate what each screen on the finished performance system program should look like. Using the rule process charts and sample screens, project staff created the performance support system. Prototype system routines were reviewed and revised to correct programming errors, to improve interface inadequacies, and to ensure user friendliness.

Activity 5: Field Testing

The performance support system shell was reviewed by selected professionals as a preliminary step to pilot testing. Data were programmed into the system to allow users to explore all possible investigative paths in the in-depth analysis. Formal and informal feedback was provided, which resulted in minor revisions in the programming (elimination of "bugs" from the system) and content of the evaluation system. The evaluation system is currently being tested within an online masters degree program in HRD at a large Midwestern university. Following this initial testing, there are plans to expand the field testing to other university level online programs and in corporate settings.

Implications and Discussion

What truly separates the vital sign concept from traditional evaluation is its use of specific outcome measures with minimal data requirements to create a cursory picture of the general status of a program. The six vital signs are measured for each individual online course. Efficiency is maximized because initial data requirements are minimal, only specific programs are examined, and only problematic vital signs are subjected to further examination.

The most obvious technical feature of this new evaluation system is its use of the electronic performance support system. Used primarily in medicine, manufacturing, and engineering up to this point, this is the first attempt to adapt advanced technology to educational evaluation. This application, of course, has its limitations. Attempting to set limits on variables in social science research always causes consternation, for fear of eliminating some possible responses from consideration. In developing investigative paths for the expert system, limits had to be set. However, the data collection instruments and procedures were carefully developed and extensively field tested. The knowledge and skill of recognized evaluation and online experts were incorporated into the system. These factors, coupled with the extensive field testing still to be completed, result in an evaluation system that is as true to human expert evaluation as is possible. It is also important to remember that the system will be operated by human evaluators and all final judgments will be made by personnel of the program online. Thus, it should not be construed that the computer is replacing the human evaluator. It is only providing valuable assistance, which will facilitate self-evaluation of online programs.

The broad applicability of the vital signs gives this system utility for online program evaluation beyond the University of Illinois. Ultimately, the model developed for use in this system, along with the computer technologies that are applied, could be adapted for use in evaluating any online course or program in the public sector as well as in private sector training and human resource development programs.

References

Alkin, M. C. (1969). Evaluation theory development. *Evaluation Comment*, 2, 2-7.

Arenson, K. W. (1998, November 2). More colleges plunging into uncharted waters of online courses. *The New York Times*.

Clark, R. E. (1994). Assessment of distance learning technology. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 63-78). Hillsdale, NJ: Erlbaum.

Draper, S. W. (1996). Observing, measuring, or evaluating courseware [On-line].
Available: http://www.psy.gla.ac.uk/~steve/Eval.HE.html

Hayes-Roth, F., Waterman, D., & Lenat, D. (1983) *Building expert systems*. Reading, MA: Addison-Wesley.

Holton, E. F. III. (1995). In search of an integrated model for HRD evaluation. In E. F. III, Holton (Ed.), *Proceedings of the 1995 Academy of Human Resource Development Annual Conference* (pp. 4-2). Baton Rouge, LA: Academy of HRD.

Jackson, G. A. (1990). Evaluating learning technology: Methods, strategies, and examples in higher education. *Journal of Higher Education, 61*(3), 294-311.

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly, 5*, 371-380.

Khan, B. H. (1997). Factors to consider when evaluating a web-based instruction course: A survey. In B. H. Khan (Ed.), *Web-based instruction* (pp. 375-378). Englewood Cliffs, NJ: Educational Technology Publications.

Kimball, L. (1998). Easier evaluation with web-based tools. *Training & Development, 52*(4), 54-55.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook* (pp. 18-1-18-27). New York: McGraw-Hill.

Kirkpatrick, D. L. (1996). Evaluation. In R. L. Craig (Ed)., *Training and development handbook* (4th ed., pp. 294-312). New York: McGraw-Hill.

Kirkpatrick, D. L. (1998). *Evaluating training programs* (2nd ed.). San Francisco: Berrrett-Koehler.

Magalhaes, M. G. M., & Schiel, D. (1997). A method for evaluation of a course delivered via the world wide web in Brazil. *The American Journal of Distance Education, 11*(2), 64-71.

McCollum, K. (1998, May 15). Accreditors are urged to prepare to evaluate distance learning. *The Chronicle of Higher Education*, A34.

Mendels, P. (1999, January 6). Universities embrace technology, but distance learning faces controversy. *The New York Times*, Education Column.

Middleton, A. J. (1997). How effective is distance education?. *International Journal of Instructional Media, 24*(2), 133-137.

Nichols, G. W. (1997). Formative evaluation of web-based instruction. In B. H. Khan (Ed.), *Web-based instruction* (pp. 369-374). Englewood Cliffs, NJ: Educational Technology Publications.

Oakes, K. (1997). The hardest question to answer about CBT. *Training & Development, 51*(9), 45-47.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: SAGE.

Pisik, G. B. (1997). Is this course instructionally sound? A guide to evaluating online training courses. *Educational Technology, 37*(4), 50-59.

Public Schools.

Raab, R., Swanson, B., Wentling, T. L., & Clark, C. (1992). Training Evaluation. Rome, Italy: Food and Agriculture Organization of the United Nations.

Ravitz, J. (1997). Evaluating learning networks: A special challenge for web-based instruction. In B. H. Khan (Ed.), *Web-based instruction* (pp. 361-368). Englewood Cliffs, NJ: Educational Technology Publications.

Raybould, B. (1995). Performance support engineering: An emerging development methodology for enabling organizational learning. *Performance Improvement Quarterly, 8*(1), 7-22.

Scriven, M. S. (1967). *The methodology of evaluation* (Perspectives of Curriculum Evaluation, and AERA monograph Series on Curriculum Evaluation, No. 1). Chicago: Rand NcNally.

Simonson, M. R. (1997). Evaluating teaching and learning at a distance. *New Directions for Teaching and Learning, 71*, 87-94.

Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record, 68*, 523-540.

Stake, R. E. (1978). The case-study method in social inquiry. *Educational Researcher, 7*, 5-8.

Stake, R. E. (1981). Setting standards for educational evaluators. *Evaluation News, 2*(2), 148-152.

Stufflebeam, D. L. (1967). The use and abuse of evaluation in Title III. *Theory Into Practice, 6*, 126-133.

Stufflebeam, D. L., Foley, W. J., Gaphart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.

Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic Evaluation*. Boston: Kluwer-Nijhoff.

Thorpe, M. (1993). *Evaluating open & distance learning* (2nd ed.). London: Longman.

Waterman, D. A. (1986). *A guide to expert systems*. Reading, MA: Addison-Wesley.

Wentling, T. L. (1980). *Evaluating occupational education and training programs*. Urbana, IL: Griffon Press.

Wentling, T. L. (1992). *Ameliorer La Qualite De La Formation*, Rome, Italy: Food and Agriculture Organization of the United Nations.

Wentling, T. L.(1993). *A trainer's guide to curriculum development*. Rome, Italy: Food and Agriculture Organization of the United Nations.

The Mixed-Method Evaluation in HRD - A Description of Methods Applied in the Industrial Development Program

Ville Nurmi
Finnish Institute of Occupational Health, Finland

This study focused on the HRD program evaluation in a case from the Finnish paper industry. Both philosophical and practical foundations of the mixed-method evaluation were described. The empirical study included description and analyses of the evaluation procedures implemented in the target program. The conclusions consisted of suggestions for the HRD practitioners, i.e., for purposes the mixed-method evaluation is useful, and what are the advantages and the disadvantages of the mixed-methodology in HRD evaluation.

Keywords: Mixed-Methods, Evaluation, HRD Program

Introduction

This study aimed to answer two questions: (1) In the light of the case analyses, what types of evaluation needs in HRD could be best served by mixed-method evaluation procedures? And, (2) what are the advantages and the disadvantages of mixed-method evaluation in HRD programs?

Evaluation in human resource development (HRD) has relied strongly on Kirkpatrick's (Kirkpatrick, 1976; Preskill, 1997) model for evaluating training programs. Regularly the evaluations which apply this model employ quantitative methodology, and the data are gathered via survey questionnaires. HRD is a complex social phenomenon to evaluate. There is no single method or methodology by which the effectiveness of development actions could be evaluated in a superior and ultimate way in order to serve the multiple needs of stakeholders. Evaluation and its methodology should be adjusted into purposes which control evaluation process (Ravishankar et al., 1997). Thus mixed-method evaluation is a more useful and purposeful design for HRD evaluation than e.g. the naturalistic or the postpositivist evaluation paradigms alone.

Mixed-method Evaluation in the Light of Philosophy of Science

The method discussion has been dominated by the two paradigms mentioned. The debate between the naturalistic or constructivist paradigm (Guba & Lincoln, 1989; Guba, 1990; Lincoln & Guba, 1985; Lincoln, 1990) and the postpositivist paradigm (Phillips, 1990; Mathiason, 1988) has emphasized the bi-polar dimensions, such as quantitative-qualitative, realism-relativism, and objectivity-subjectivity. The contructivists, like Guba and Lincoln, are proponents of the purist stance toward mixed-method evaluation; they argue that different paradigms embody completely different assumptions about human nature and the world. According to their understanding, it is not possible to mix different inquiry paradigms within a single study. (Greene et al., 1989; Greene & Caracelli, 1997.)

The postpositivist school is not as rigid as the constructivist school. According to this paradigm, there are philosophical differences between the paradigms of social sciences. But the philosophical assumptions of different paradigms are logically independent, thus paradigms can be mixed. Quantitative methods are found useful in studying social phenomena, even though it is not believed that human behavior could be controlled and defined by a set of natural laws, as logical-positivists believe. The postpositivists try to identify consistencies affecting human behavior, but not by denying the usefulness of qualitative methodology. The third stance toward mixed-method evaluation is the dialectical position in the methodology toward the mixed-method debate. Its idea is that philosophical differences between paradigms exist, and that they are important. Differences should be honored and used both within and across studies in order to achieve enhanced understanding. (Greene et al., 1989; Greene & Caracelli, 1997.)

In this study these paradigms were seen as complementary choices rather than competing methodological schools. The empirical study part employed both the postpositivist and the constructivistic paradigms in the evaluation of the HRD program. The general methodological discourse has shifted its focus in the mixed-method direction, too.

The Mixed-Method Evaluation in Practice

Copyright © 1999 Ville Nurmi

The mixed-method in evaluation means that multiple data gathering or analyzing procedures are combined and used within the same study. Different procedures could be used to answer the same evaluation questions, or they could be employed in answering different evaluation questions. The mixed-methodology in HRD evaluation as in other social science evaluation fields too, is applied in order to generate deeper knowledge and broader insight into the phenomenon under analysis. All methods have limitations and biases, and multiple methods can be helpful in counteracting those biases. (Datta, 1997; Greene & Caracelli, 1997; Waysman & Savaya, 1997.)

Triangulation was the first well known attempt to mix methods in social sciences. Psychologists in the 1950s used multiple methods in studies on personal traits to strengthen the validity of their studies. In triangulation, two or more methods are used to assess the same phenomenon in order to strengthen validity, but also to seek convergence of the findings. *Bracketing* is the second multiple method design. It seeks estimates on the correct answer. *Complementarity* is the third type of mixed-method evaluation design. Its fundamental idea, is to reach a deeper understanding of the evaluation object. There are three kinds of complementarity evaluations: different methods are applied (1) to evaluate different study components or phenomena, (2) to assess the plausibility of validity threats, or (3) to enhance the interpretability of the assessments of a phenomenon. (Greene et al., 1989.)

In addition to mixing methods or methodologies within the same study, it is possible and in some cases purposeful to mix data sources, investigators, and theories. Data source triangulation means that multiple data sources are employed in a single evaluation. Investigator triangulation means that several, at least two, researchers analyze the same object either using different methods, or validating the findings produced by one method. Triangulating theories are also possible and acceptable in evaluation studies. The theoretical frame of reference of evaluation studies are eclectic in nature; thus not much attention is paid to theory triangulation. The most discussed type of triangulation is mixing methodologies or methods. (Denzin, 1978; Mathison, 1988.)

A Brief Description of the HRD Program

The present study was focused on a HRD program in the Finnish paper industry in 1995-1996. Metsä-Serla corporation invested USD 500 million into the world's biggest fine paper mill. That time Metsä-Serla was the fifth biggest pulp and paper producer in Europe. It has production units all around the European continent, and its products are sold in Asia, South America, and North America, besides Europe.

Within the technical paper mill construction project there was a HRD program; the new technology required modern and up-to-date vocational qualifications from the workers. Over 150 workers were hired to the new fine paper mill through careful recruitment procedures. The HRD program lasted 8 months including classroom courses, practical training periods, excursions, simulations, and some organized leisure time activities, too. One third of the trainees were without previous work experience in the paper industry. This made HRD one of the key factors in terms of the successful start-up of the new paper mill. (Nurmi, 1998.)

The HRD program was one of the biggest ever conducted in the Finnish private sector with its USD 10 million budget. In addition to its size, there were other innovative aspects, too. Behavioral scientists were hired to plan, implement, and evaluate the HRD activities, while the HRD personnel regularly include only technical experts. Stakeholder involvement in the planning and evaluation processes, the great emphasis given to the training in so called soft skills, high educational requirements for hired workers, and finally the multiple evaluation procedures implemented both during and after the HRD program, were crucial meaningful innovations in the HRD program of Metsä-Serla.

The HRD program was conducted simultaneously with the technical paper mill construction project. The final goal of both projects was the successful start-up of the new paper mill in August 1996 (Figure 1). There were also other objectives set for HRD, which were related to the start-up goal. In addition to training for the workers HRD included management training and training of the maintenance personnel. This study focused on the training of the operators, i.e. workers of the new paper production line (Nurmi, 1998).

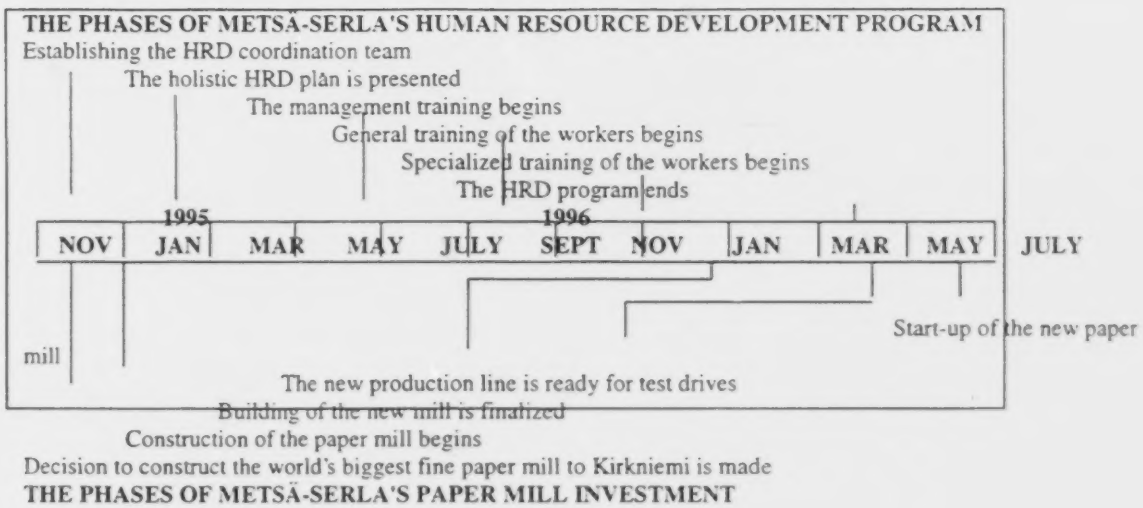


Figure 1. The chronological description of Metsä-Serla's paper mill investment including the human resource development program.

The Empirical Study

The empirical part of the study describes the evaluation methods that were used in the target program. The description focused on the following points: What were the evaluation questions? What was the purpose of the evaluation? What data gathering methods were applied? What data analysis procedures were applied? Was the evaluation formative or summative? The method description formed a basis for the analysis of the advantages and the disadvantages of the evaluation procedures.

The Mixed-Method Application in this Study

The postpositivist frame of reference was used in this study. However the evaluation procedures analyzed in this study contained naturalistically oriented elements, too. Postpositivism means that by applying quantitative surveys and multivariate analysis, it is possibile to identify general patterns which affect learning and satisfaction. This study represents complementarity in terms of the mixed-method design, rather than triangulation or bracketing. Different methods were used to assess different components of the phenomenon, which in this case was the effectiveness of the HRD program. The final purpose of mixing methods was to reach deeper understanding, than that gained via a single method.

Different data sources were used also in this study. This is evident always when the aim of the evaluation is to analyze the effectiveness of the whole program, instead of a single aspect. Data sources are described in the results section. Even though one researcher conducted further analyses on the effectiveness of the HRD program, other investigators conducted data gathering and analyzing of procedures during the program. Therefore this study also employed investigator triangulation.

The Description of Evaluation Procedures Applied in the HRD Program

The first evaluation question, i.e. "How did the trainees feel after each course?", represented the traditional first level evaluation of Kirkpatrick (1976) which emphasizes the reactions of the trainees. The purpose of this phase was to improve the program by revising course contents and other activities in a direction that would be most beneficial for the trainees. The data were gathered by a survey including numerical and open-ended questions. The names of the respondents were asked; this offered an opportunity to later discuss with the trainees their comments and suggestions. The data were analyzed immediately after each training period to get the formative evaluation information as quickly as possible. Means and standard deviations were produced. If the mean of the course with

most of the items was under 3.5 (the scale 1-5; 1 failure, 5 success), some improvements or further information was required. Qualitative data were not analyzed for research purposes, but information from the open-ended answers was used in practice (Table 1).

The second question, "How did the trainees progress and learn?", was based on Kirkpatrick's second level evaluation (Kirkpatrick, 1976), where the focus is on learning in terms of knowledge, skills, and attitudes. Attitudes were not assessed in this case, however. Learning was analyzed for both formative and summative purposes. The formative purpose was the same as with the first evaluation question. The summative evaluation was conducted for the needs of the formal educational system; trainees were offered an opportunity to pursue a vocational certification within the development program. The vocational system, controlled and coordinated by the Finnish National Board of Education, required documentation of the trainees' progress. Learning was evaluated by multiple-choice and open-ended questions. Besides these knowledge-oriented measurements, there were both manual and automated practical simulations which assessed the skills of the trainees in a specific competence area. (Brown, 1996; Knight, 1995.) The results of the learning assessments were coded into grades for the purposes of the certification system. Descriptive statistics were produced to evaluate the success of the trainees, and to compare different trainee populations (Table 1).

Based on the learning assessments, further data analyses were conducted to identify the latent factors effecting the trainees' success and behavior. This summative type of evaluation was done by multivariate data analysis. The results of the assessments were first summarized into a few groups by factor analysis; these factors were then predicted in multiple regression by demographics. The findings will be used when new HRD actions are planned and implemented in the future (Table 1).

The trainees' feelings and experiences were evaluated also more in depth than merely by reaction surveys. Ten trainees were asked to be informants, i.e. monitor persons. Their assignment was to keep a learning journal during the eight-month program, and to be interviewed three times: at the beginning, in the middle, and at the end of the program. The purpose of this evaluation was to obtain understanding of the trainees' behavior and actions in the program, but this naturalistic data could also be used in the planning process of future programs. The data analyses at this phase have not yet been completed. The data analyses will include case study analyses in which each monitor person represents one single case, and there will be also a phenomenological section for defining concepts, such as vocational competence in the paper industry (Table 1).

The experiences of the HRD personnel were gathered and analyzed for the same purposes as the trainees' experiences, but the data gathering methods were different. The HRD experts wrote their comments on a qualitative questionnaire, which was used instead of interviews to save time through avoiding time-consuming interview transcription. The management was interviewed to gather their beliefs and experiences from the program. Even though the top managers did not attend the development actions on a daily basis, their experiences of the merit of the program will be crucial, because future decisions to invest resources on HRD will be done by these managers. At the end of the program, future needs assessment was carried out. The supervisors discussed the program with every worker, and together they filled out a needs assessment form (Table 1).

The last evaluation question, "Were the organizational HRD goals attained?", included four subquestions, because there were four organizational goals set for HRD. The attainment of each goal was analyzed in order to evaluate the effectiveness of the program. Multiple data gathering and analyzing methods were employed at this phase of the evaluation. (Nurmi, 1998; Table 1.) This evaluation had a bottom-line orientation, thus representing Kirkpatrick's (1976) fourth evaluation stage, where the focus is on the training effects on an organization's results.

Trainees' reactions

This evaluation was beneficial, because the results were gathered and analyzed immediately after the training and the results were available a couple days after the course. The reaction questionnaire was easy to answer, and the open-ended questions offered an opportunity for the respondents to explain their reactions more deeply than in the numerical sections. The use of real names brought more validity; the respondents were responsible for their comments and stood behind their views. The mainly statistical evaluation data could be used as a source for generalizations about the reactions of the trainees, and different courses of the program could be compared in terms of trainee satisfaction.

On the other hand, some disadvantages were found, too. Because the program consisted of several courses, the trainees had to fill out the same questionnaire many times. This was perceived frustrating, and sometimes led the subjects to fill out forms mechanically, without reflecting their real feelings. Sometimes, a minor incident at a course, e.g. one negative comment by a trainer, received much more attention than would have been appropriate. And this can greatly affect the reaction evaluations, although there were not necessarily big problems in the course.

Finally, it is a costly investment to transfer reaction data into computer files, in such a format that they can be used in further statistical analyses.

Table 1

Evaluation Procedures of the HRD Program of Metsä-Serla Corporation

EVALUATION QUESTION	PURPOSE	DATA GATHERING	DATA ANALYSIS	SUMMATIVE/ FORMATIVE
How did the trainees feel after each course?	to clarify reactions of trainees, in order to improve the program	survey questionnaire after each course	descriptive statistics: means & standard deviations	formative
How did the trainees progress and learn?	to identify areas needing more emphasis	multiple-choice, open-ended questions; simulations	grades ⇒ descriptive statistics, mean comparisons	formative
--- " ---	to identify trainees who obtained a vocational certification	--- " ---	--- " ---	summative
Which demographics predicted the trainees' success?	to identify fundamental factors effecting the trainees success	--- " ---	multivariate data analysis	summative
What were the trainees' feelings and experiences in the program?	to understand the trainees actions, and to improve future HRD	interviews and learning journals	qualitative data analysis: case study & phenomenology	summative
What were the HRD personnel's experiences after the program?	to improve future programs	qualitative questionnaire	case study	summative
How did the management evaluate the HRD program?	to understand the management's attitudes, which affect future human investment decisions	interviews	case study	summative
What future development needs did trainees have?	to coordinate future HRD actions in the organization	questionnaires filled with supervisors	content analysis	formative
Were the organizational HRD goals attained?	to define the merit of the HRD program	production information, interviews, assessments	qualitative comparisons, content analysis, analysis of variance	summative

Advantages and Disadvantages of Evaluation Procedures

Assessments of learning

This evaluation phase offers an instrument for instructional designers and trainers to follow up the progress of the trainees. If there are problems, e.g. learning difficulties either within the whole trainee population or individually, HRD actions can be revised, and in addition, individual solutions could be produced to make learning and studying more meaningful and effective. Learning assessments are useful for trainees also because they give information about their own progress. The trainees can improve their study habits and put more emphasis on areas where they have learning difficulties. Quantitative assessment data can be used for further statistical analyses, which

have potential for generalizations within the population, but across to the other similar populations, too. Simulations can be used-as one type of learning measurement, because the trainees can test their skills in a real-life setting, which is usually risk-free, unlike their job assignments in the production line.

It is difficult to find relevant learning assessments in industrial HRD. Most of the assessments have been developed case by case, and this is time-consuming. It requires also considerable resources to give grades in paper-and-pencil examinations. In the field, there is a lack of comparative assessment results. In the case of Metsä-Serla, there was one assessment of the conceptual mastery of paper production (Leppänen, 1993, Leppänen & al., 1996), in which the data had been gathered from other similar organizations. But all the other evaluation tools were developed in the program. If there are no control groups or comparative populations, the question arises, what is the purpose of learning assessment. Multivariate data analyses require large sample sizes, and their results are not necessarily useful for a single company. These types of analyses are rather conducted by research institutes for scientific purposes. The trainees may feel that learning assessment is used simply to control their actions, and as a result their focus may change from learning for work to learning for examinations, as in the case of many formal educational settings.

Evaluation of Experiences

Naturalistic evaluation procedures produced more depth data about the experiences, feelings and beliefs of stakeholders than reaction evaluation or learning assessments. This information is a necessary foundation for a deeper understanding of the HRD program. Qualitative data describes those nuances and thinking patterns which are regularly left out of surveys. It would be more meaningful for the respondents to explain and reflect upon their experiences, than simply fill out questionnaires.

The collection and analysis of naturalistically oriented, qualitative data from the participants of the HRD program and from other stakeholders is a huge investment in terms of time and money. The designing of data gathering procedures also requires more resources than the planning of questionnaire surveys, because the same interview structure can not be used in different cases. Like all qualitative research, also this kind of evaluation is not a valid base for generalizations. The definition of the final merit of the program does not often employ qualitative type of data as the major data source.

NeedsAssessment

It was felt that more detailed answers could be received, if the needs are assessed during the HRD program. If this evaluation had been conducted after the start-up of the paper mill, the workers as well as the supervisors, would have felt that they have more important things to think about and do than the HRD needs assessment. On the other hand, the needs assessment was conducted by questionnaire, and this might have led the supervisors and their subordinates to fill out the form more carefully. Thus the data gathering method might have caused lack of interaction and communication between the supervisors and the workers.

The Attainment Analysis of Organizational Goals

This evaluation was beneficial in defining the final merit of the program. It is relevant for the organization to receive information on how the organizational goals were attained, if the stakeholders, including the management, have been involved in the goal-setting process. Regularly organizational goals are related to bottom-line results, thus this type of evaluation could be used in assessing the financial effects of human investments.

A political orientation is inherent in this evaluation procedure; it is crucial importance to know whose goals are important and defined valuable. Even though an organizational perspective is needed for the bottom-line focus, in many cases the top-management, which has to set the goals, does not have any connection to the HRD practice. It is difficult to define and evaluate the effects of HRD on the financial results of organizations, but this is needed in order to legitimize the development actions and secure beneficial investment decisions in the future.

Conclusions

Mixed-methods in evaluating HRD programs is a purposeful design, when the aim is to assess a large program from multiple viewpoints. Conducting an evaluation according to Kirkpatrick's (1976) model requires both quantitative and qualitative procedures in order to be informative and satisfactory for every stakeholder group, and successful in

documenting the varying aspects of the program. Even the first level evaluation, i.e. reaction evaluation, produces deeper data, if also open-ended qualitative questions are applied.

The usefulness of the mixed-methods depends on the stakeholder group that is served by the evaluation. Sometimes purely quantitative evaluation procedures could produce the required information to analyze the financial effects of HRD, even though it is almost impossible, separately from other organizational effects to show unquestionable evidence of the economic benefits of successful HRD actions. Evaluation data on the experiences and feelings of stakeholders, such as the personnel, the management, HRD experts, external trainers, and even shareholders, could be presented within the organization in a way that is interesting and meaningful to the rest of the work community.

It should not be expected that mixed-method evaluation produces results that will automatically become an integrated whole. There may be inconsistencies, or even controversies between the results produced by different evaluation methods. This should be taken as evidence of the richness of social phenomena, and evaluators should try to understand why there are differences, and whether there is a reason for differences to appear on certain occasions. (Mathison, 1988.) The results of this study showed that there are several advantages and disadvantages in applying mixed-method evaluation. The mixed-method approach requires more expertise, more time, more money, and more implementers than evaluation conducted by a single method. Evaluation needs should be analyzed case by case, and possible benefits of mixed-methods should be weighed, before the final evaluation is designed (Waysman & Savaya, 1997).

This case study offered suggestions for HRD practitioners who are in the role of evaluator, too. Evaluation should be well planned before HRD implementation takes place. In this case, the evaluation was mainly planned during the training, thus preassessments could not be made. If financial effects of the HRD program are required as a result of the evaluation, then expected financial outcomes should be defined with the managers of the target organization (Swanson & Mattson, 1997). Otherwise it is quite difficult to isolate the effects of HRD after the program implementation has taken place. In gathering qualitative data, evaluators should be careful not to produce large amounts of information, which is neither needed nor useful. If the evaluation is conducted by a team of evaluators, instead of a single evaluator, there should be mutual agreement about the responsibilities and assignments of the team members. In the Metsä-Serla case, there was not a special evaluation expert; the personnel of the HRD team were also responsible on the evaluation procedures. Thus, there were no extra expenses from the program evaluation; HRD coordinators designed and implemented the evaluation, in addition to their main responsibilities in the program. But a HRD expert does not necessarily have the qualifications of an evaluator. A person who plans and implements evaluation in practice obviously is not an evaluator. The same difference exists between a laboratory technician and an analytical chemist, the former implements measurements, as does the latter, but only the latter designs research studies and interprets study findings (Scriven, 1996).

Finally, conclusions can be drawn on some ideas about the relationship between quantitative and qualitative paradigms in social sciences. Those paradigms are not alternatives for each other; qualitative can not replace quantitative, or vice versa. The paradigms seek answers for different questions. Therefore it is not recommendable to name some research areas or objects to be suitable just for one research design. There is always data between the researcher and the study object. It is not a piece of the object, even though some qualitatively oriented researchers assume so. For both survey and qualitative studies, data are based on human speech, and there is no base for an argument that qualitative data are somehow closer to the object than survey data. Quantity and quality are not the opposite, they are different concepts for organizing and understanding reality. (Töttö, 1997.) As mentioned at the beginning of this paper, the method debate and evaluation practices should have shifted their emphasis from the quantitative-qualitative contradiction to a more open, but at the same time analytical and critical way of mixing different methods in evaluation research.

References

Brown, G. (1996). Assessing student learning in higher education. London: Routledge.

Datta, L. E. (1997). A pragmatic basis for mixed-method designs. In J. C. Greene, & V. J. Caracelli (Eds.), Advances in mixed-method evaluation: The challenges and benefits of the integrating diverse paradigms (pp. 33-46). San Francisco: Jossey-Bass.

Denzin, N. K. (1978). Triangulation. In N. K. Denzin (Ed.), The research act: An introduction to sociological methods. New York: McGraw-Hill.

Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. In J. C. Greene, & V. J. Caracelli (Eds.), Advances in mixed-method evaluation: The challenges and benefits of the integrating diverse paradigms (pp. 5-18). San Francisco: Jossey-Bass.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 11(3), 255-274.

Guba, E. G., & Lincoln, Y. S (1989). Fourth generation evaluation. Newbury Park: Sage.

Guba, E.G. (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), The paradigm dialog (pp. 17-30). Newbury Park: Sage.

Kirkpatrick, D. L. (1976). Techniques for evaluating training programs. Training and Development Journal, 30(15), 3-17.

Knight, P. (1995). Assessment for learning in higher education. London: Kogan Page.

Lincoln,, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Newbury Park: Sage.

Lincoln, Y.,S. (1990). The making of a contructivist: A remembrance of transformation past. In E. G. Guba (Ed.), The paradigm dialog (pp. 67-87). Newbury Park: Sage.

Phillips, D. C. (1990). Postpositivistic science: Myths and realities. In E. G. Guba (Ed.), The paradigm dialog (pp. 31-45). Newbury Park: Sage.

Mathison, S. (1988). Why triangulate. Educational Researcher, 17(3), 13-17.

Nurmi, V. (1998). Koulutuksen tavoitteiden saavuttaminen organisaatiomuutoksessa - tapaus paperiteollisuudesta. [The Attainment of HRD Goals in the Organizational Change Process- A Case Study from the Paper Industry]. Työ ja Ihminen, 12(4).

Preskill, H. (1997). HRD evaluation as the catalyst for organizational learning. In E. Holton, III (Ed.), Conference Proceedings of Academy of Human Resource Development , 49-56.

Ravishankar, L. & Miller, Z. (1997). A qualitative and quantitative approach to evaluating leadership training. In E.Holton, III (Ed.), Conference Proceedings of Academy of Human Resource Development, 57-63.

Scriven, M. (1996). Types of evaluation and types of evaluator. Evaluation Practice, 17(2), 151-161.

Swanson, R. A., & Mattson B. W. (1997). Development and Evaluation of the Critical Outcome Technique. In E.Holton, III (Ed.) Conference Proceedings of Academy of Human Resource Development, 64-71.

Töttö, P. (1997). Pirullinen positivismi: Kysymyksiä laadulliselle tutkimukselle. [The Devilish Positivism: Issues for Qualitative Research Paradigm]. Jyväskylä: Yliopistopaino.

Waysmann, M., & Savaya, R. (1997). Mixed-methods evaluation: a case study. Evaluation Practice, 18(3), 227-237.

The Effect of Two Training Systems on Trainee Satisfaction, Learning, Job Behavior, Retention and Organizational Performance

Maria Hruby Moore
The Ohio State University

The effect of two systems of training using the same course content with two different groups of trainees revealed there was no significant difference between trainee satisfaction, learning, or individual job performance outcomes. However there were significant differences between learning ratings assessed by the corporate instructor versus the manager, as well as the retention rate and organizational performance measures for each training system. Implications for HRD professionals and the use of distance technology are discussed.

Keywords: Training Systems, Learner Outcomes

This evaluation research study focused on the effectiveness of two training systems for delivering the same sales training content. Training System One (1) is described as a manager-led, self-paced videotape delivery system. Training System Two (2) was different in that it used the same corporate instructor with a group method over two-way interactive videoconferencing. This study focused on comparing training effectiveness outcomes for the two training systems using Kirpatrick's (1996) Four Level Evaluation Model as the general format. The same training goals and objectives were used by both systems. Objectives defined what the trainee was expected to do as a result of the training related to job performance. Within the three domains of learning: cognitive (thinking), affective (feeling) and psychomotor (task performance), the sales training was expected to help participants integrate these three domains and demonstrate their skills throughout the course. The primary instructional techniques used were structured discussions, videotapes to demonstrate behavior, role-plays, written exercises, and practice. A large amount of interaction is required by the training objectives.

The outcomes expected of the sales training was to help new representatives learn the job and become productive contributors earlier. By providing structured on-the-job training (Jacobs & Jones, 1995) at the office location, the context and conditions under which the trainee ultimately performed was an essential component of both training systems. The difference was that one system was conducted independently with the manager at a negotiated pace, while the other used a scheduled group method using distance technology with a corporate trainer. Acceptable standards of job performance were emphasized in both systems.

Theoretical Framework

Issues regarding interactivity, course structure and cognitive style were explored and measured in this study. Interactivity is the active participation in the instruction/training process. Wagner (1998) suggests that performance improvement specialists should look at the outcomes of interactions: Do they change the learner and move them toward action to the stated goal? Interactivity is also referred to as "dialog" which is described by Moore and Kearsley (1996) as any interplay of words, actions, ideas and other interactions between the teacher and the learner when one gives instruction and the other responds. It is measured in this study by the total number of hours recorded by the trainee based on their interaction with the manager or instructor throughout their training period. Research literature in cognitive and perceptual psychology supports the importance of interaction in the learning process (Neisser, 1976 and Gardner, 1985).

By taking a systems view, course structure (Moore & Kearsley, 1996) refers to the course design variables that consist of elements such as learning objectives, content themes, information presentation, case studies, pictures, illustrations, exercises, projects and tests. In this study, Training System 1 is less structured than Training System 2, since more variability and modification is tolerated in Training System 1. Moore identifies both dialog (interaction) and structure in his theory of transactional distance (Moore, 1973; 1983). He suggests that it is these two variables that can be manipulated to reduce the amount of psychological distance between a learner and the instructor when engaged in a distance learning-teaching process regarding interest in a common content. Distance technology can be used to increase or decrease the amount of structure and dialog in the teaching-learning transaction. In this case, distance technology was applied by the Training Department to Training System 1 to create Training System 2.

Copyright © 1999 Maria Hruby Moore

Cognitive style is a distinctive and habitual manner that an individual uses to organize and process information. According to Witkin and Goodenough (1981), there are two defined cognitive styles: field dependence and field independence. Field dependent people use a global, holistic approach to learning while field independents use an analytical, linear approach to learning and problem solving. Cognitive style is a psychological construct identified and validated by the work of Witkin and his colleagues (Goodenough, 1976). It was measured in this study by their instrument, the Group Embedded Figures Test (GEFT) (Witkin, et. al., 1971). Research on cognitive style suggests that field dependent learners tend to be socially oriented, prefer collaboration, are female, and are extrinsically motivated. Field independent learners tend to be analytical, rely on self-structured situations, prefer competition, are male, and intrinsically motivated (Witkin, et. al., Shih, et. al., 1998).

Research Problem and Questions

The major issue of concern in this evaluation research study was: *Which delivery system is more effective in achieving the training as well as the organization's objectives?* This is a difficult problem for any field research, were there is little experimental control over internal and external threats to validity. Likewise, the participants in this study were only those who were still one the job after one year of going through the training. It was important to see if these two groups differed significantly on certain characteristics. Only limited data was available on those that had exited the organization. The major research questions were:

1. What are the characteristics of the two groups on the following measures: age, gender, years of experience (work, sales and industry), education, amount of interaction, and cognitive style?
2. Is there a difference in the overall satisfaction with the content and quality of the training between the two groups?
3. Is there a difference in learning as rated by both the trainees and the instructor/manager between the two groups?
4. Is there a difference in the behavior or individual performance as a result of the training between the two groups?
5. Is there a difference in the total performance and retention rates between these two groups?

Research Design

The research design was a static group comparison, which is considered a pre-experimental design by Campbell and Stanley (1963). Again, the Kirpatrick's (1996) Four Levels of Evaluation format was adopted for this study.

Population

Participants who were trained in the first half of 1997 were the population selected. Managers were asked to secure participation from the trainee in their office for this study. 78 out of 119 trainees in the manager-led training and 136 out of 408 trained in the corporate instructor-led system actually participated in this study. This time frame was selected since it had the same corporate instructor across all classes; and little change in the instructional content was introduced into either system. The data collected were only from trainees still with the organization after one year of employment. Table 1 describes the characteristic differences between the two groups.

First, there were significantly more men than women trainees. Training Group 1 had 87 or 64% men compared to 49 or 36% women. Training Group 2 was more evenly distributed, but still had 45 or 58% men and 43 or 42% women. Second, the descriptive statistics noted in Table 1 above indicate there are some other differences between the groups. Training Group 1 appears to have less sales and industry experience, more education and lower GEFT scores than Training Group 2. Statistical significance at the .05 level would indicate the two groups differ on sales and industry experience, education, and GEFT scores even though they do not differ significantly on age or overall years of work experience. Overall, there was no significant difference in the hours of interaction recorded by either group. A score of 12 or higher tends toward field independence and the reliability coefficient of the GEFT is .82 (Witkin, Oltman, Raskin, Karp, 1971).

However, the following comparison group results suggest that there were no significant differences between the two groups of trainees on satisfaction with the training program, learning as perceived by the trainee, and job behavior. Differences between the two groups were statistically significant at the .05 level for learning as

perceived by the corporate instructor versus the managers' ratings. Also, there were differences between retention rates and overall performance for the two groups.

Description of the Trainee Population

Variable	Mean TS1	Std. Dev.	Mean TS2	Std.Dev.	T-test	Sig. (2-tailed)
Age (#years old)	41.8	9.1	39.4	10.4	1.56	.120
# Years Experience						
Work	16.8	8.1	14.7	10.4	1.51	.132
Sales	4.6	6.4	7.1	7.7	-2.43	.016 *
Industry	1.9	3.1	3.4	5.7	-2.03	.043 *
# Years Education	16.5	1.7	15.1	2.7	3.92	.000 *
# Hrs. Interaction	30.3	29.1	24.2	27.6	1.54	.126
GEFT Score	11.3	4.4	13.3	4.8	-2.91	.004 *

* Notes Significance at $\alpha = .05$

Table 1: Comparison of Means for Characteristics of the Two Groups Training Under System 1 versus 2

Results and Findings

Level One: Reaction => Satisfaction with Training Program

Reaction refers to the trainees' perception of the worth of the training activity (Kirkpatrick, 1996). Training participant reactions were recorded as their satisfaction with the content, quality, and delivery of the training. Reactions were measured in this evaluation as a total score on the BAET Evaluation form. The evaluation contained both quantitative and qualitative measures. The quantitative measures were 15 items, rated with a Likert scale of 1-5, which focused on the satisfaction. The reliability coefficient alpha for the BAET calculated by SPSS was .86.

Statistical results indicated at the .05 level that the two groups did not differ significantly on their overall reaction to their training experience. Those trained by managers using videotapes had a mean satisfaction score of 55.28 versus a 55.97 for those trained by the corporate instructor under videoconferencing. See Table 2 for the comparison.

Training Satisfaction Score

Training System	Mean	Std.Dev.	T-test	Sig. (2-tailed)
1. Manager-led, Videotape	55.28	10.31	-.381	.704
2. Corporate Instructor-led, Videoconferencing	55.97	11.31		

Table 2: Comparison of Trainees' Satisfaction Score in Training System 1 and 2

Level Two: Learning => Self and Manager vs. Instructor Ratings

Learning refers to the acquisition of the necessary knowledge and skills to perform the job. For the job of an account executive, conceptual and process knowledge has to be understood. Second, knowledge and demonstration of sales and computer skill has to be acquired. A variety of factors effect whether the trainee can or will learn the material presented in training. These factors include the learner's experience, motivation, aptitude and attitude toward what is to be learned. The work environment and the appropriate tools can also effect learning outcomes. Learning was measured using the trainee's self-rating on the BAET Evaluation and the instructor's or

managers' rating using the same scale. Both ratings are considered subjective measures. The BAET Likert rating scale was:

1=Poor *2=Fair* *3=Good* *4=Very Good* *5=Excellent*

Results indicated at the .05 level that there were no significant differences between the trainees' average learning rating of 3.81 versus 3.65 for Training System 1 compared to 2. Again, this average represents how the trainees' perceived and assessed their own learning. However, there was a significant difference at the .05 level in how the managers and the corporate instructor assessed their trainees' learning. The average rating by the managers in Training System 1 was 2.33 and the average rating by the corporate instructor was 1.72 for trainees in Training System 2. See Table 3 for the comparison. Speculation for the differences in manager and corporate instructor ratings is discussed in the Conclusion.

Trainees' and Manager vs. Instructor Measures of Learning Outcomes

Training System	Learning Rating by	Mean	Std.Dev.	T-test	Sig. (2-tailed)
1. Manager-led, Videotape	Trainee	3.81	.82	1.39	.167
	Manager	2.33	.63	5.81	.000*
2. Corporate, Video-conferencing	Trainee	3.65	.81		
	Instructor	1.72	.63		

* Notes Significance at $\alpha = .05$

Table 3: Comparison of Trainees' and Managers' vs. Corporate Instructor's Learning Ratings in System 1 and 2

Level Three: Job Behavior => Actual Performance on the Job After One Year

Job behavior refers to the degree to which the knowledge and skills learned in the training transferred to the job. In this study, performance on the job was measured as the total number of placements and dollar billings after one year on the job. These are the two individual job performance measures expected of an account executive.

Results again indicated that there was no significant difference at the .05 level between the average number of placements and the average dollars billed between the two groups. Those who went through Training System 1 had an average of 6.2 placements and billed an average of $71,635 while those in Training System 2 had an average of 5.5 placements and billed an average of $76,389. See Table 4 for the comparison.

Actual Job Behavior as Measured by Placements and Billings

Training System	Job Behavior	Mean	Std.Dev.	T-test	Sig.
1. Manager-led, Videotape	Placements #	6.19	5.19	.952	.342
	Billings $	$71,635	$ 59,037	-.306	.760
2. Corporate, Video-conferencing	Placements #	5.48	6.69		
	Billings $	$76,389	$152,405		

Table 4: Comparison of Trainees' Job Behavior in Training System 1 and 2

Level Four: Organization Results => Retention and Collective Organizational Impact

Organization results refer to the degree to which a training effort supports business objectives. Organization results are the reason for using training. Account executives perform a critical role for the

organization. Their collective competence as professionals has lead to the company being identified as the industry leader. Although difficult to isolate training as the critical factor, and to measure all organization performance outcomes, some effort to capture impact is important. In this study, results were measured by the total placements and billings over one year (1997-98) for Training System 1 and 2. Those trained by managers with video (n=119) produced 631 placements and billed $7,306,744.00. Those trained by the corporate instructor over videoconferencing (n=409) produced 1,370.5 placements and billed $19,097,316.

A per desk average, or PDA statistic, was also calculated to determine which training system performed better overall. PDA is a corporate, year-end statistic currently used by the company to measure and compare office performance. In this case, it was used to measure and compare training system performance. It is calculated as the total placements and dollars billed divided by those remaining with the organization at the end of the year. Those trained by the manager-led system had a PDA of 8.1 placements and $93,676.21 in billings. Those trained by the corporate system had a PDA of 10.1 placements and $140,421.44 in billings. In this study, Training System 2 performed 33% higher in comparing the PDAs of the two groups.

Retention is another measure used to assess impact on an organization. A retention rate for each training system was calculated. Retention was the total number of account executives who remained on the job after one year and participated in the study divided by the total number of participants that went through the training system. Results indicated a rate of 66% for Training System 1 and 33% for Training System 2. See Table 5 for a comparison of total placements, billings, per desk averages, and retention rates for the two groups.

Overall Organizational Results

Training System	n of N	Total # Placements and $ Billed	Per Desk Avg. (PDA)	% Retention
1. Manager-led, Videotape	78 of 138	631.0 - $ 7,306,744	8.1 - $ 93,676	66%
2. Corporate-led, Video Conferencing	136 of 409	1,370.5 - $19,097,316	10.1 - $140,421	33%

Table 5: Comparison of Trainees' Total Placements and Billings with Retention Rates in Training System 1 and 2

Conclusion

Both systems of training are effective in meeting the training and organizational goals of the company. The data suggests that both training systems are equally effective at achieving trainee reaction, learning, and job performance objectives from the perspectives of the trainees. However, in comparing the managers' and the corporate trainer's learning evaluation scores of the trainees, there was a significant difference. Likewise, at the organizational level we observe the two training systems contributing to organizational performance, but in different ways. Both are effective at eliciting different outcomes that are important to the organization. Training System 1 excels in retention and Training System 2 trains more people, produces more placements, higher dollar billings, and higher PDAs.

Which delivery system is <u>more</u> effective in achieving the training as well as the organization's objectives?

Based on the data collected from the first three levels of this evaluation, both groups of trainees' reactions or satisfaction, learning, and behavior measures all appear to support that trainees were satisfied with their training experience and understood and performed the training objectives. Although both training systems were able to meet the training objectives, differences in learning ratings by the managers versus the corporate instructor were observed. Also, overall organizational results and performance outcomes were different as noted in total placements and billings, per desk average results, and retention rates.

To address the difference in learning rating by the corporate instructor versus the managers, perhaps managers that have to train new hires are more careful in their selection process, and want to hold on to marginal or average performers for the long-run. It is also possible that these managers have a more difficult time letting a

trainee go who is at least producing something, due to the personal investment the manager made in training them. On the other hand, the corporate instructor is not able to develop a personal relationship with the trainer as the manager. Maybe the distance between the corporate instructor and trainee allows him to be more objective. Therefore, his average rating of trainees' could reflect why there is a lower retention rate. It is possible that managers who send their trainees through Training System 2 are not as careful in their selection process or as involved in the training as they should be. It should also be noted that the corporate instructor does not have control over the work environment.

On side of Training System 2, the production from those trained resulted collectively in more total placements and higher billings. Training System 2 is simply able to train more people. Although retention is lower, the training allows some participants to produce something before they exit the firm. Another speculation is that the lower retention rate may be a sign that any participant who is not a "good match" can make a decision to exit the business much more easily in Training System 2. Perhaps the training style, or distance of the instructor to the trainee, also effects retention. If the training style of the corporate instructor was not congruent with the cognitive style of the learners, then more trainees may leave. There are also a number of factors that can explain the retention rate that are not related to training. These include the individual owner's and/or manager's strategy, compensation structure, the management practices, the individual office environment and/or the trainees personal goals and motivation. Supervised feedback and personal involvement of the manager in training is critical in both systems.

Implications for HRD and Notes for Further Research

Evaluation contains the word "value" in it. Values stated in this organization's mission statement are: *to be a global network ... and the world's preferred and pre-eminent provider of....* Currently they are the industry leader. Both training systems support company's mission; however, Training System 2 may have more benefits. Since "value" can be defined as merit in both economic terms and character qualities, both training systems show no significant difference between trainees' average satisfaction, learning, dollar billings and number of placements. Economic value is achieved by account executives performing their jobs. The company's character can be measured by the trainees' satisfaction with the training and their learning outcomes that lead to professionalism in the field. Ultimately, individual performance is the outcome that leads to desired organization results. Both training systems support the development of account executives who can provide their customers with the solutions they desire.

In comparing the two training systems, effectiveness has been achieved by both systems without any significant difference at the first three levels. However, when looking at organizational impact, new questions are raised. *Is training more participants who don't stay, good for business in the long run? Which training system is more efficient in terms of actual costs and financial benefits?* The concept of efficiency could be explored further to find out which delivery system is less costly to the organization. This issue can also be looked at in terms of the actual costs of the manager's lost revenue generating time compared to the instructor's time training plus the cost of videoconferencing technology.

Another interesting outcome appeared in the trainees' qualitative evaluation comments. From those involved in the videoconferencing system, almost all trainees' responded positively to the question "Do you feel the videoconferencing technology helped or hindered your learning?" In addition, of the 136 responses, 56 mentioned the ability for "more interactions" with other offices and 42 described the ability to better develop a "network". This issue should be explored further as part of the company's mission is to be a "Global Network". *Which system better supports building customer and office relationships - and a global network?*

Likewise, given the overall gender and cognitive style of trainee participants that are retained by both systems tends toward males and field independence. Perhaps a sales job, such as this, is male dominated and requires field independence to be successful in the training and the job. Research suggests that the greater the fit between the training system and the learner's style, the more effective training is likely to be (O'Conner, Bronner, & Delaney, 1996). Currently, most trainees do not have the choice as to which system they are trained under. The issue of cognitive style matched to training system will be explored further.

With respect to distance technology, more needs to be understood about interaction, course structure and cognitive style for HRD practitioners and researchers to understand how individual and organizational performance can be improved with training and other organizational development interventions. The development of distance and communication technologies spurred by the Internet and videoconferencing are taking off (Imel, 1998; Schlosser and Anderson, 1994). Weinstein (1997) suggests that the appeal of these technologies is the interactivity of the teacher and students and their ability to share text, graphics, audio, video, and virtual reality experiences despite physical separation. In this study, interaction was required by the training objective and the amount of

interactivity was not statistically different for either system. There was a difference in the way the training was structured and the cognitive styles of the two groups. Finally, further exploration of the qualitative comments from this study may shed light on the concept of network and relationship building and the impact of these on a sales job. The important issue is how distance technology can support individual performance on the job and organizational outcomes. This study begins to suggests that with careful attention to training objectives, interaction, course structure, and cognitive style, that distance technology can be employed to deliver individual and organizational performance.

References

Campbell, D.T. and Stanley, J.C. (1963). Experimental and quasi-experimental design for research. Dallas, TX: Houghton Mifflin Company.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Goodenough, D.R. (1976). The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin*, 83, 675-694.

Kirpatrick, D.L.(1996). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler Publishers.

Jacobs, R.L., and Jones, M. (1995). *Structured on-the-job training: Unleashing employee expertise in the workplace*. San Francisco, CA: Berrett-Koehler Publishers.

Imel, S. (1998). *Distance Education*. ERIC Myths and Realities. http://www.ericacve.org.

Moore, M.G. (1973). Towards a theory of independent learning and teaching. *Journal of higher education*. (44), 661-679.

Moore, M.G. (1983). On a theory of independent study. In D. Seward, D. Keegan, & B. Holmberg (Eds.), *Distance education: International Perspectives*. London: Croom Helm.

Moore, and Kearsley, (1996). *Distance education: A systems view*. Belmont, CA: Wadsworth Publishing.

Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: W.H. Freeman.

O'Conner, B.N., Bronner, M.B., & Delaney, C.D. (1996). *Training for Organizations*. Cincinnati, OH: South-Western Educational Publishing.

Schlosser, C.A. and Anderson, M.L. (1994). Distance education: Review of literature. Washington, D.C.: Association for educational communications and technology, (ED 382 159).

Shih, C-C., Ingebritsen, T., Pleasants, J., Flickinger, K, and Brown, G. (1998). Learning strategies and other factors influencing achievement via web courses. *Proceedings from the 14th Annual Conference on Distance Teaching and Learning*. University of Wisconsin: Madison, Wisconsin.

Wagner, E. D. (1998). Interaction strategies for online training designs. *Proceedings from the 14th Annual Conference on Distance Teaching and Learning*. University of Wisconsin: Madison, Wisconsin.

Weinstein, P. (1997, May-June). Education goes the distance. *Technology and learning*. 17(8), 24-25.

Witkin, H.A, and Goodenough, D.R., (1981). *Cognitive styles: Essence and origins. Field Dependence and field independence*. Psychological Issues. Monagraph 51, New York: International Universities Press, Inc.

Witkin, H.A., Oltman, P.K., Raskin, E. and Karp, K.A. (1971). *A manual for the embedded figures test*. Consulting Psychologists Press, Inc.